

# Estimasi Fungsi Regresi Dalam Model Regresi Nonparametrik Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel

<sup>1</sup>Budi Lestari, <sup>2</sup>Nur Chamidah, dan <sup>3</sup>Toha Saifudin

## Abstract

Regression model of bi-respond nonparametric is a regression model which is illustrating of the connection pattern between respond variable and one or more predictor variables, where between first respond and second respond have correlation each other. In this paper, we discuss the estimating functions of regression in regression model of bi-respond nonparametric by using different two estimation techniques, namely, smoothing spline and kernel. This study showed that for using smoothing spline and kernel, the estimator function of regression which has been obtained in observation is a regression linier. In addition, both estimators that are obtained from those two techniques are systematically only different on smoothing matrices.

**Keywords:** kernel estimator, smoothing spline estimator, regression function, bi-respond nonparametric regression model.

## Abstrak

Model regresi nonparametrik birespon adalah suatu model regresi yang menggambarkan pola hubungan antara dua variabel respon dan satu atau beberapa variabel prediktor dimana antara respon pertama dan respon kedua berkorelasi. Dalam makalah ini dibahas estimasi fungsi regresi dalam model regresi nonparametrik birespon menggunakan dua teknik estimasi yang berbeda, yaitu smoothing spline dan kernel. Hasil studi ini menunjukkan bahwa, baik menggunakan smoothing spline maupun menggunakan kernel, estimator fungsi regresi yang didapatkan merupakan fungsi linier dalam observasi. Selain itu, kedua estimator fungsi regresi yang didapatkan dari kedua teknik estimasi tersebut secara matematis hanya dibedakan oleh matriks penghalusnya.

**Kata Kunci :** Estimator Kernel, Estimator Smoothing Spline, Fungsi Regresi, Model Regresi Nonparametrik Birespon.

## 1. Pendahuluan

Misal diberikan pasangan data  $(y_{ki}, t_{ki})$ ,  $k=1,2$ ;  $i=1,2,\dots,n_k$  mengikuti model regresi nonparametrik birespon sebagai berikut:

$$y_{ki} = f_k(t_{ki}) + \varepsilon_{ki}, \quad a_k \leq t_k \leq b_k \quad (1)$$

dengan  $k$  adalah menyatakan banyaknya respon, dan bentuk fungsi regresi  $f_1$  dan  $f_2$  tidak diketahui dan diasumsikan mulus termuat dalam ruang Sobolev  $W_2^m[a_k, b_k]$ , serta  $\varepsilon_{ki}$  ( $k=1,2$ ;  $i=1,2,\dots,n_k$ ) adalah error random independent dengan mean nol dan

<sup>1</sup>Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jember

Jalan Kalimantan 37 Kampus Tegal Boto, Jember 68121, Indonesia.

<sup>1</sup>E-Mail: budi.fmipa@unej.ac.id

<sup>2,3</sup>Departemen Matematika, Fakultas Sains dan Teknologi,

Universitas Airlangga, Surabaya,

Jalan Mulyorejo Kampus C UNAIR, Surabaya 60115 Indonesia.

variansi  $\sigma_{ki}^2$  (Lestari *et al.*, 2012). Tujuan mendasar dari regresi nonparametrik birespon adalah untuk mengestimasi fungsi regresi yang tidak diketahui,  $f_k \in W_2^m[a_k, b_k]$ , dalam model (1).

Dalam makalah ini diperhatikan kembali model regresi nonparametrik birespon yang dibahas oleh Wang *et al.*, (2000). Selanjutnya berdasarkan model tersebut, dibahas estimasi fungsi regresi berdasarkan dua estimator berbeda, yakni estimator smoothing spline dan estimator kernel. Chamidah dan Lestari (2016) menggunakan pendekatan estimator spline untuk mengestimasi fungsi regresi, tetapi untuk kasus variansi sama (homoskedastik). Meskipun Aydin (2007) telah membahas tentang dua teknik estimasi fungsi regresi, yakni estimator smoothing spline dan estimator kernel, tetapi Aydin (2007) menggunakan estimator smoothing spline and estimator kernel tersebut hanya untuk mengestimasi fungsi regresi dalam model regresi nonparametrik unirespon saja. Jadi Aydin (2007) belum membahas kedua teknik estimasi fungsi regresi tersebut untuk model regresi nonparametrik birespon. Sedangkan Wang *et al.*, (2000) belum membahas estimator kernel untuk mengestimasi fungsi regresi dalam model regresi nonparametric birespon. Oleh karena itu, dalam makalah ini akan dibahas dua teknik estimasi fungsi regresi, yakni estimator smoothing spline dan estimator kernel, dalam model regresi nonparametrik birespon untuk kasus variansi tidak sama (yakni,  $\sigma_{ki}^2$ ,  $i = 1, 2, \dots, n_k$ ;  $i = 1, 2$ ). Pembahasan estimasi fungsi regresi dalam model (1) menggunakan estimator smoothing spline dan estimator kernel berturut-turut disajikan dalam bab 2 dan bab 3.

## 2. Estimasi Fungsi Regresi Berdasarkan Estimator Smoothing Spline.

Misalkan  $\underline{y} = (y_1, y_2)'$ ,  $\underline{f} = (f_1, f_2)'$ ,  $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2)'$ , dan  $\underline{t} = (t_1, t_2)'$  dengan  $\underline{y}_k = (y_{k1}, y_{k2}, \dots, y_{kn_k})'$ ,  $\underline{f}_k = (f_k(t_{k1}), f_k(t_{k2}), \dots, f_k(t_{kn_k}))'$ ,  $\underline{\varepsilon}_k = (\varepsilon_{k1}, \varepsilon_{k2}, \dots, \varepsilon_{kn_k})'$ ,  $\underline{t}_k = (t_{k1}, t_{k2}, \dots, t_{kn_k})'$ ,  $k = 1, 2$ ;  $i = 1, 2, \dots, n_k$ . Oleh karena itu, model (1) dapat dinyatakan dalam notasi vektor sebagai berikut:

$$\underline{y} = \underline{f} + \underline{\varepsilon} \quad (2)$$

dengan  $E(\underline{\varepsilon}) = \underline{0}$ , dan  $Cov(\underline{\varepsilon}) = [W(\underline{\sigma}^2)]^{-1} = diag(W_1(\underline{\sigma}_1^2), W_2(\underline{\sigma}_2^2))$  (Lestari *et al.* (2012), dan Lestari *et al.* (2017)). Estimasi fungsi  $\underline{f}$  dalam model (2) berdasarkan estimator smoothing spline diperoleh dari penyelesaian masalah meminimumkan penalized weighted least-square (PWLS), yakni menentukan  $\hat{\underline{f}}$  yang meminimumkan PWLS:

$$\text{Min}_{\underline{f}_1, \underline{f}_2 \in W_2} \{ (\sum_{k=1}^2 n_k)^{-1} (y_1 - \underline{f}_1)' W_1 (y_1 - \underline{f}_1) + (y_2 - \underline{f}_2)' W_2 (y_2 - \underline{f}_2) + \lambda_1 \int_{a_1}^{b_1} (f_1^{(2)}(t))^2 dt + \lambda_2 \int_{a_2}^{b_2} (f_2^{(2)}(t))^2 dt \} \quad (3)$$

untuk nilai  $\underline{\lambda} = (\lambda_1, \lambda_2)' > \underline{0}$ . Komponen pertama dalam (3) menyatakan jumlah kuadrat residual yang merupakan ukuran *goodness of fit*, sedangkan komponen dalam (3) yang diboboti oleh  $\underline{\lambda}$ , menyatakan ukuran kemulusan (*smoothness*) estimasi fungsi regresi  $\underline{f}$ .

Nilai  $\lambda_k$  ( $k = 1, 2$ ) dalam (3) dikenal sebagai parameter penghalus (*smoothing parameter*). Jika nilai  $\lambda_k$  bervariasi dari 0 sampai  $+\infty$ , maka solusi juga bervariasi dari interpolasi ke suatu model linier. Untuk  $\lambda_k \rightarrow +\infty$ , penalty kemulusan mendominasi dalam (3), dan estimasi spline mendekati konstan. Untuk  $\lambda_k \rightarrow 0$ , penalty kemulusan tidak muncul dalam (3), dan estimasi spline menginterpolasi data. Jadi parameter

penghalus  $\lambda_k$  memainkan peranan penting dalam mengontrol *trade-off* () antara *goodness of fit* yang dinyatakan oleh:

$$\left(\sum_{k=1}^2 n_k\right)^{-1} (\underline{y}_1 - \underline{f}_1)' W_1 (\underline{y}_1 - \underline{f}_1) + (\underline{y}_2 - \underline{f}_2)' W_2 (\underline{y}_2 - \underline{f}_2)$$

dan kemulusan estimasi diukur dengan:

$$\lambda_1 \int_{a_1}^{b_1} (f_1^{(2)}(t))^2 dt + \lambda_2 \int_{a_2}^{b_2} (f_2^{(2)}(t))^2 dt.$$

Selesaian berdasarkan smoothing spline untuk masalah meminimumkan (3) adalah dikenal sebagai “*natural cubic spline*” dengan knots  $t_1, t_2, \dots, t_{n_k}$  ( $k = 1, 2$ ). Bertolak dari hal ini, suatu interpolasi spline terstruktur khusus yang bergantung pada suatu nilai yang dipilih  $\lambda_k$  merupakan suatu pendekatan yang sesuai untuk fungsi-fungsi  $f_k$  ( $k = 1, 2$ ) dalam model regresi nonparametric birespon (1). Selanjutnya, misalkan  $\underline{f} = (\underline{f}_1, \underline{f}_2)'$  dengan  $\underline{f}_k = (f_k(t_{k1}), f_k(t_{k2}), \dots, f_k(t_{kn_k}))'$ ,  $k = 1, 2$  adalah vektor dari nilai-nilai fungsi  $f_k$  ( $k = 1, 2$ ) pada titik-titik knot  $t_1, t_2, \dots, t_{n_k}$  ( $k = 1, 2$ ). Estimator smoothing spline  $\hat{f}_\lambda$  dari vektor atau nilai estimasi untuk data  $\underline{y} = (\underline{y}_1, \underline{y}_2)'$  dengan  $\underline{y}_k = (y_{k1}, y_{k2}, \dots, y_{kn_k})'$ ,  $k = 1, 2$  dapat diperoleh dari Lestari *et al.* (2012), dan Lestari *et al.* (2017):

$$\hat{f}_\lambda = \hat{\underline{y}} = \begin{pmatrix} \hat{f}_{1,\lambda_1} \\ \hat{f}_{2,\lambda_2} \end{pmatrix} = H(\underline{\lambda}) \underline{y} \quad (4)$$

dengan  $\hat{f}_\lambda$  adalah suatu *natural cubic spline* dengan knot-knot pada  $t_1, t_2, \dots, t_{n_k}$  ( $k = 1, 2$ ) untuk suatu parameter penghalus  $\underline{\lambda} = (\lambda_1, \lambda_2)' > \underline{0}$ , dan  $H(\underline{\lambda})$  dikenal sebagai matriks penghalus definit-positif (simetri) yang bergantung pada  $\underline{\lambda}$  dan titik-titik knot  $t_1, t_2, \dots, t_{n_k}$  ( $k = 1, 2$ ), tetapi tidak bergantung pada  $\underline{y}$ . Untuk referensi umum tentang smoothing spline, dapat ditemukan pada Wahba (1990), Green dan Silverman (1994), dan Eubank (1999).

### 3. Estimasi Fungsi Regresi Berdasarkan Estimator Kernel.

Filosofi utama dari model regresi nonparametrik adalah untuk mengestimasi fungsi regresi  $f$  menggunakan suatu rata-rata terbobot dari data mentah dimana bobot-bobotnya merupakan suatu fungsi dari jarak dalam ruang- $t$ . Secara khusus, bobot tersebut adalah suatu fungsi turun dari jarak. Untuk model regresi nonparametrik unirespon, suatu skema pembobotan dari tipe ini diberikan oleh Nadaraya-Watson (1964) yakni bobot yang berkaitan dengan observasi  $y_j$ , untuk prediksi di  $t_i$  diberikan oleh persamaan sebagai berikut:

$$w_{ij} = \frac{K\left(\frac{t_i - t_j}{h}\right)}{\sum_{j=1}^n K\left(\frac{t_i - t_j}{h}\right)} = \frac{K(u)}{\sum_{j=1}^n K(u)} \quad (5)$$

dengan  $K(u)$  adalah suatu fungsi turun dari  $u$ , dan  $h > 0$  disebut *bandwidth* atau parameter penghalus.  $K(u)$ , fungsi kernel yang didapatkan merupakan fungsi kepadatan probabilitas, misal Gaussian. Fungsi kernel tersebut adalah simetris (Yatchew, 2003 dan Wand & Jones, 1995).

Selanjutnya, berdasarkan persamaan (5) dan dengan memperhatikan model regresi nonparametrik birespon yang diberikan dalam (1), maka bobot berkaitan dengan observasi dari respon ke- $k$ ,  $y_{kj}$ , untuk prediksi di  $t_{ki}$  diberikan oleh:

$$w_{(k)ij} = \frac{K_k \left( \frac{t_{ki} - t_{kj}}{h_k} \right)}{\sum_{j=1}^{n_k} K_k \left( \frac{t_{ki} - t_{kj}}{h_k} \right)} = \frac{K_k(u)}{\sum_{j=1}^{n_k} K_k(u)}, \quad k=1,2. \quad (6)$$

Berdasarkan pada persamaan (6), diperoleh estimasi kernel dari fungsi  $f$  dalam model regresi nonparametrik birespon (1) pada suatu titik pengepasan  $t_{ki}$  sebagai berikut:

$$\hat{f}_{\tilde{k}}(t_{ki}) = \hat{y}_{ki} = \sum_{j=1}^{n_k} w_{(k)ij} y_{kj} = \tilde{w}'_{kj} \tilde{y}_k, \quad i=1,2,\dots,n_k, \quad k=1,2. \quad (7)$$

Setiap  $n$  titik-titik data diberikan suatu bobot berbeda  $w_{(k)ij}$ ,  $j=1,2,\dots,n_k$  untuk sembarang titik pengepasan  $t_{ki}$ . Dalam notasi matriks, persamaan (7) dapat dinyatakan sebagai berikut:

$$\hat{f}_{\tilde{k}} = W \tilde{y} \quad (8)$$

dengan

$$W = \text{diag} (W_{k1}, W_{k2}, \dots, W_{kn_k}), \text{ and}$$

$$W_{ki} = \begin{pmatrix} w_{(k)11} & w_{(k)12} & \cdots & w_{(k)1n_k} \\ w_{(k)21} & w_{(k)22} & \cdots & w_{(k)2n_k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{(k)n_k1} & w_{(k)n_k2} & \cdots & w_{(k)n_k n_k} \end{pmatrix}, \quad k=1,2.$$

Matrix  $W$  menotasikan *kernel hat* atau matriks penghalus kernel. Seperti halnya pada *ordinary least square* (OLS) dimana *hat matrix* digunakan untuk mentransformasikan  $y_j$  ke  $\hat{y}_i$ , *hat matrix* kernel digunakan untuk mentransformasikan  $y_j$  ke  $\hat{y}_i$ . Prediksi kernel pada suatu titik,  $t_{ki}$ , diperoleh dengan menggunakan persamaan (7), dan dengan penggantian “ $ki$ ” dengan “ $k1$ ”. Jadi prediksi kernel pada suatu titik  $t_{ki}$  dapat ditulis sebagai berikut:

$$\hat{f}_{\tilde{k}}(t_{k1}) = \tilde{w}'_{k1} \tilde{y}_k = \left( w_{(k)11}, w_{(k)12}, \dots, w_{(k)1n_k} \right) \begin{pmatrix} y_{k1} \\ y_{k2} \\ \vdots \\ y_{kn_k} \end{pmatrix}, \quad k=1,2. \quad (9)$$

Akhirnya, dengan cara sama seperti pada estimasi smoothing spline dalam persamaan (4), dan dengan memperhatikan persamaan (8) dan (9), maka estimasi fungsi regresi dari model regresi nonparametrik birespon dalam (1) berdasarkan estimator kernel diberikan oleh:

$$\hat{\tilde{f}} = \begin{pmatrix} \hat{f}_1(t_1) \\ \hat{f}_2(t_2) \end{pmatrix} = \begin{pmatrix} W_{k1} & 0 & \cdots & 0 \\ 0 & W_{k2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{kn_k} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad k=1,2. \quad (10)$$

#### 4. Kesimpulan

Berdasarkan persamaan (4) dan (10), estimator smoothing spline ( $\hat{f}_{\lambda}$ ), dan estimator kernel ( $\hat{f}$ ) adalah estimator-estimator yang linier dalam observasi, dan kedua estimator tersebut adalah bias untuk  $f$ . Hasil estimasi yang diperoleh dengan menggunakan estimator smoothing spline and estimator kernel secara tampilan formula atau persamaan memiliki kemiripan, sedangkan secara matematis keduanya hanya dibedakan oleh matriks-matriks penghalusnya, yakni oleh suatu matriks penghalus definit-positif (simetri) yang bergantung pada  $\lambda$  dan pada titik-titik knot untuk pendekatan estimator smoothing spline, dan suatu *hat matrix* kernel atau matriks penghalus kernel untuk pendekatan estimator kernel.

#### Ucapan Terima Kasih

Makalah ini merupakan sebagian dari hasil penelitian PDUPT (Penelitian Dasar Unggulan Perguruan Tinggi) Lanjutan (Tahun ke-2 dari rencana 2 tahun). Untuk itu kami mengucapkan terima kasih kepada Direktur DRPM-DIKTI yang telah mendanai penelitian ini melalui Hibah Penelitian PDUPT (Lanjutan) Tahun Anggaran 2018.

#### Daftar Pustaka

- [1]. Aydin, D., 2007. A comparison of the nonparametric regression models using smoothing spline and kernel regression. *World Acad. Sci. Eng. Technol.*, **26**, 730-734.
- [2]. Chamidah, N. dan Lestari, B., 2016. Spline Estimator in Homoscedastic Multi-Response Nonparametric Regression Model in Case of Unbalanced Number of Observations. *Far East Journal of Mathematical Sciences (FJMS)*, **100**(9), 1433-1453.
- [3]. Eubank, R.L., 1999. *Nonparametric Regression and Smoothing Spline*, Marcel Dekker, New York.
- [4]. Green, P.J. dan Silverman, B.W. 1994. *Nonparametric Regression and Generalized Linear Models*. Chapman Hall. New York.
- [5]. Hardle, W., 1991. *Applied Nonparametric Regression*. Cambridge University Press. Cambridge.
- [6]. Lestari, B., Budiantara, I.N., Sunaryo, S., dan Mashuri, M., 2012. Spline smoothing for multi-response nonparametric regression model in case of heteroscedasticity of variance. *J. Math. and Stat.* **8**, 377-384.
- [7]. Lestari, B., Fatmawati, dan Budiantara, I. N., 2017. Estimasi Fungsi Regresi Nonparametrik Multirespon Menggunakan Reproducing Kernel Hilbert Space Berdasarkan Estimator Smoothing Spline. *Proceeding of National Seminar on Mathematics and Its Applications (SNMA) 2017*, Airlangga University, Surabaya, pp. 243-250.
- [8]. Nadaraya, E. A., 1964. On Estimating Regression. *Theory Prob. Applications*, **10**, 186-190.

- [9]. Schimek, M. G., 2000. *Smoothing and Regression*. John Wiley & Sons, New York.
- [10]. Wand, M. P. dan Jones, M.C., 1995. *Kernel Smoothing*. Chapman Hall, New York.
- [11]. Wang, Y., W. Guo dan Brown, W.B., 2000. Spline smoothing for bivariate data with applications to association between hormones. *Staistica Sinica*. **10**, 377-397.
- [12]. Watson, G. S., 1964. Smooth Regression Analysis. *Sankhya Series A*. **26**, 359-372.
- [13]. Yatchew, A., 2003. *Semiparametric Regression for the Applied Econometrician*. Cambridge University Press, Cambridge.